AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 104:167–176 (1997)

Y-Chromosome Specific Alleles and Haplotypes in European and Asian Populations: Linkage Disequilibrium and Geographic Diversity

R.J. MITCHELL,* L. EARL, AND B. FRICKE
School of Genetics and Human Variation, La Trobe University, Bundoora, Melbourne, Victoria, Australia 3083

KEY WORDS Y chromosome; RFLPs; haplotypes; 92R7; pYαI; p12f; p21A1; migration; Italians; Greeks; Slavs; Asians

ABSTRACT Variation on the Y chromosome may permit our understanding the evolution of the human paternal lineage and male gene flow. This study reports upon the distribution and non random association of alleles at four Y-chromosome specific loci in four populations, three Caucasoid (Italian, Greek and Slav) and one Asian. The markers include insertion/deletion (p12f), point mutation (92R7 and pYαI), and repeat sequence (p21A1) polymorphisms. Our data confirm that the p12f/*Taq*I 8 kb allele is a Caucasoid marker and that Asians are monomorphic at three of the loci (p12f, 92R7, and pYαI). The alleles at 92R7 and pYαI were found to be in complete disequilibrium in Europeans. Y-haplotype diversity was highly significant between Asians and all three European groups ($P < 0.001$), but the Greeks and Italians were also significantly different with respect to some alleles and haplotypes ($P < 0.02$). We find strong evidence that the p12f/*Taq*I 8 kb allele may have arisen only once, as a deletion event, and, additionally, that the present-day frequency distribution of Y chromosomes carrying the p12f/8 kb allele suggests that it may have been spread by colonising sea-faring peoples from the Near East, possibly the Phoenicians, rather than by expansion of Neolithic farmers into continental Europe. The p12f deletion is the key marker of a unique Y chromosome, found only in Caucasians to date, labelled 'Mediterranean' and this further increases the level of Y-chromosome diversity seen among Caucasoids when compared to the other major population groups. Am J Phys Anthropol 104:167–176, 1997.

The human Y chromosome is unique among nuclear chromosomes because it is inherited solely by males in the hemizygous state and, with the exception of the pseudoautosomal region (PAR) at the tip of the short arm of Y, does not undergo recombination at meiosis. A second, smaller, PAR on Yq has been reported by Freije et al. (1992). Therefore, the Y chromosome behaves as a haploid, and every Y chromosome retains a record of the mutational events that occurred on its ancestors. It is for this reason that the Y chromosome may become as important to the study of male lineages as the mitochondrial chromosome (mtDNA) has come to be for female lineages (Cann et al., 1987; Brega et al., 1986).

That the Y chromosome has not yet proven to be as useful as mtDNA for studying sex-specific lineages occurs principally because of the shortage of sequence polymorphism on it (Ngo and Lucotte, 1986; Jakubiczka et al., 1989; Malaspina et al., 1990;

*Correspondence to: Dr. R.J. Mitchell, School of Genetics and Human Variation, La Trobe University, Bundoora, Victoria, Australia 3083. E-mail: genrjm@genome.gen.latrobe.edu.au

Received 27 February 1996; accepted 25 March 1997.

Spurdle and Jenkins 1992a; Dorit et al., 1995; Whitfield et al., 1995). It is unknown why the Y-specific chromosome is deficient in sequence variation. The opposite might be expected given the origin of the chromosome (Charlesworth, 1991). One reason may be the smaller effective size of Y chromosomes compared with that for autosomes which are three times larger. An alternative explanation relates to the lack of recombination on the Y chromosome. It has been argued that if selection occurred for a favourable mutation on the non-recombining part of the Y chromosome it would result in fixation of a single Y haplotype. The favoured allele would spread through the population accompanied by the rest of the particular male specific chromosome on which the mutational event occurred (Dorit et al., 1995).

Compounding the difficulties associated with the shortage of markers is that some of the Y-specific polymorphisms reported 1) have a complex structural basis (Torroni et al., 1990; Spurdle and Jenkins, 1992b; Spurdle and Jenkins, 1993a); 2) are more variable than was initially reported (Mathias et al., 1994); 3) have multiple independent origins for alleles (Spurdle et al., 1994a; Hammer and Horai, 1995); and 4) have, at present, an unidentifiable ancestral state, since the polymorphism is absent in other higher primates (Mathias et al., 1994).

Despite these limitations, the few available Y-chromosome specific restriction fragment length polymorphisms (RFLPs) are of considerable importance for studies of microevolution, especially those of historical migration patterns and sex-specific gene flow (Maynard Smith, 1990). They are also valuable in measuring population affinities (Torroni et al., 1990; Mitchell et al., 1993; Spurdle and Jenkins, 1993a; Spurdle et al., 1994a; Hammer, 1994; Hammer and Horai, 1995), genetic admixture (Spurdle et al., 1994b; Torroni et al., 1994) and Y-chromosome evolution (Mathias et al., 1994; Jobling, 1994; Seielstad et al., 1994; Hammer, 1994, 1995; Jobling and Tyler-Smith, 1995). The majority of these studies have focussed upon Caucasoid populations and, to a lesser extent, Africans. To date very little has been reported for Asian groups.

Four Y-chromosome specific RFLPs merit further attention. Probe 92R7, an *Eco*RI–*Eco*RI fragment of about 2.1 kb subcloned from cY92 (Cooke et al., 1985), comprises a moderately repeated sequence which detects seven bands in a *Hin*dIII digest of male DNA, but none in females (Mathias et al., 1994). One of these *Hin*dIII fragments is polymorphic, and can be either 4.6 kb (allele 1) or 6.7 kb (allele 2) in size. As no polymorphism was detected with eight other restriction enzymes, Mathias et al. (1994) concluded the polymorphism is due to a point mutation.

The centromeric alphoid satellite DNA locus consists of tandemly repeated ~170 bp subunits organised into 5.7- or 6.0-kb units which are hypervariable (Tyler-Smith and Brown, 1987). In addition to its hypervariable array size (Oakey and Tyler-Smith, 1990; Mathias et al., 1994), pYαI, a 6 kb alphoid unit from the alphoid satellite DNA locus (DYZ3), detects a polymorphism, due to a point mutation, which is seen as the presence (allele 1) or absence (allele 2) of 6.0 kb repeat units containing internal *Hin*dIII sites (producing 4.1 and 1.9 kb fragments) in addition to the 5.7 kb repeat units. It has been suggested that the 6.0 kb alphoid units are ancestral to the 5.7 kb units (Oakey and Tyler-Smith, 1990).

Probe p12f (DYS11) detects a *Taq*I or *Eco*RI RFLP in which the most common allele is represented by a 10 kb fragment. In a variable number of males this 10 kb fragment (allele 1) is absent and the 8 kb fragment has a correspondingly increased intensity (allele 2). The polymorphism is suggested to result from an insertion-deletion event (Casanova et al., 1985).

Probe p21A1 (DYZ8) detects a *Taq*I RFLP in a repeat sequence that most probably arose from the loss of a *Taq*I site in a short alternating repeat unit of 4 kb and 7 kb, thus generating a 11 kb fragment (Jakubiczka et al., 1989). The 4 kb and 7 kb fragments (allele 1) represent the regular repeat, and the 4, 7 and 11 kb the mutant pattern (allele 2). Because this RFLP occurs in a repeat sequence the possibility of multiple independent origins of the mutant allele, in both place and time, has been recognised (Jakubiczka et al., 1989).

This study reports on these four Y-chromosome specific RFLPs in three European groups and a sample of Asians to determine 1) if the loci are polymorphic in each, 2) the extent of non random association between alleles at the loci in each group, and 3) the extent of intra- and inter-population variability observed among alleles and haplotypes. This analysis involves combining the present data with those on p12f and p21A1 in the same subjects reported in Mitchell et al. (1993). Comparisons are also made between the present findings and the limited data available from other populations.

MATERIALS AND METHODS

Population samples

Ninety-two Italian-born and 63 Greek-born males who were participants in a study of 'ethnicity, diet and disease' conducted in Melbourne each donated a blood sample for DNA analysis. All subjects had to complete a detailed questionnaire and undergo an extensive interview in their respective native language before they were allowed to participate in the study, and accordingly these samples are very well defined. Birthplace information was used to allocate subjects to particular geographic divisions of each country. Italy was divided into three regions, north-central, south, and Sicily, but the smaller Greek sample permitted subdivision into the mainland and islands only. The latter included Crete and Cyprus as well as the Aegean and Ionian islands.

Twenty-four Asian and 17 Slavic males, all resident in Melbourne, were selected by surname analysis, a methodology which produces a less well defined sample. The Asians were principally from Vietnam and China (including Hong Kong), while the Slavs were drawn from the states of the former Yugoslavia. No intra-population analysis was performed on these two samples.

DNA analysis and RFLP detection

Genomic DNA was extracted, digested, electrophoresed and transferred to nylon membranes using standard techniques (Sambrook et al., 1989). The hybridisation conditions and protocols followed with probes p12f and p21A1 are given in Mitchell et al. (1993) and those with pYαI and 92R7 are described in Mathias et al. (1994). Following hybridisation and washing, the fragments were visualised by autoradiography at −70°C with Fuji film.

Statistical analysis

Non-random association of alleles are measured by Fisher's Exact Test or the G test. The significance of the heterogeneity of Y alleles and haplotypes across populations is tested by the G test.

RESULTS

Allele frequencies

92R7 (No HGM locus assigned). The numbers and frequencies of each allele observed in the regional and total samples of the four groups are given in Table 1. Asians, all of whom have allele 1 (4.6 kb), are significantly different from all European groups (Asians vs. Italians $P < 0.01$; Asians vs. Greeks and Asians vs. Slavs $P < 0.05$). In Italy the frequency of allele 1 appears to vary inversely with latitude, rising from a low of 52% in the north-central area to a maximum of 68% in Sicily, with the south displaying an intermediate frequency (58%). Among Greeks, those born in the islands display a considerably higher frequency of allele 1. Slavs are more similar to Greeks than to Italians. The difference between Greeks and Italians is significant ($P < 0.01$).

pYαI (HGM locus DYZ3). Asians are monomorphic, all carrying allele 1, and this distinguishes them from the three European groups (Asians vs. Italians $P < 0.01$; Asians vs. Greeks and Asians vs. Slavs $P < 0.05$) (Table 1). The frequency of alleles in Italy varies inversely with latitude and the difference between Greeks and Italians is significant ($P < 0.01$).

p12f (HGM locus DYS11). The frequency of the 8 kb allele (allele 2) in Asians and Slavs is given in Table 1 together with the data on Italians and Greeks from Mitchell et al. (1993). The frequency in Slavs is intermediate between that seen in Greeks and Italians. Only one Asian possessed allele 2, and the possibility of Caucasoid admixture cannot be excluded. None of the pairwise com-

TABLE 1. Distribution of alleles at four Y-specific RFLP loci in Italians, Greeks, Slavs and Asians

Population and birthplace	Probe/Enzyme											
	92R7/*Hin*dIII allele 1 (4.6 kb)			pYαI/*Hin*dIII allele 1 (6.0 kb +)			p12f/*Taq*I allele 2 (8 kb)[2]			p21A1/*Taq*I allele 2 (−)[2]		
	n	No.	%	n	No.	%	n	No.	%	n	No.	%
Italy												
North-central	31	16	51.6	35	18	51.4	46	4	8.7	46	31	67.4
South	19	11	57.9	19	11	57.9	21	9	42.9	22	17	77.3
Sicily	22	15	68.2	24	16	66.7	27	7	25.9	27	18	66.7
Total[1]	75	43	57.3	81	46	56.8	97	20	20.6	98	68	69.4
Greece												
Mainland	46	36	78.3	48	38	79.2	47	14	29.8	49	34	69.4
Islands	16	14	87.5	17	14	82.4	15	7	46.7	16	11	68.7
Total	62	50	80.7	65	52	80.0	62	21	33.9	65	45	69.2
Slavs	17	12	70.6	17	12	70.6	17	4	23.5	11	5	45.5
Total Europe	154	105	68.2	163	110	67.5	176	45	25.6	174	118	67.8
Asians	24	24	100.0	24	24	100.0	24	1	4.2	23	20	87.0

[1]Includes three individuals who gave no birthplace information.
[2]Data for Italy and Greece taken from Mitchell et al. (1993).

parisons among the European groups is statistically significant, and Asians were significantly different from Greeks only ($P < 0.01$).

Probe p21A1 (HGM locus DYZ8). This locus is polymorphic in all four populations, with the frequency of allele 2 ranging between 45% in Slavs and 87% in Asians. None of the differences is significant.

Linkage disequilibrium

There is complete linkage disequilibrium between the alleles of 92R7 and pYαI in Europeans, with allele 1 of 92R7 found, without exception, with allele 1 of pYαI (haplotype 11), and, conversely, allele 2 of 92R7 is always found with allele 2 of pYαI (haplotype 22). All Asians have haplotype 11. Examination of Table 3 in Mathias et al. (1994) indicates that this disequilibrium is also present in their Caucasian sample. While alleles at 12f and 21A1 are in random association (Mitchell et al., 1993), each is in significant disequilibrium with alleles at 92R7/pYαI ($P < 0.01$).

Haplotype distributions

Alleles were combined into 1) two three-loci Y haplotypes, p12f/92R7/pYαI because these alleles possibly represent single mutational events and p21A1/92R7/pYαI because there is evidence of iso-allelism at p21A1), and 2) one four-loci Y haplotype, and the distributions of each were examined.

Haplotype 12f, 92R7 and pYαI. Because of complete disequilibrium between 92R7 and pYαI, four haplotypes are expected in Europeans but only three were observed (Table 2). Allele 2 (8 kb) at the p12f locus has a relatively high frequency, but is found only in one background, a Y chromosome containing allele 1 at the 92R7 locus and allele 1 at pYαI (haplotype 211). Allele 1 (10 kb) at the p12f locus, however, is found in both possible combinations. These data strongly suggest that the p12f polymorphism resulted from a deletion event.

Though the allelic associations are constant across the four populations, the frequencies of haplotypes exhibit considerable variation. The Asian population is significantly different from all European groups ($P < 0.001$). However, Greeks and Italians are also different from each other ($P < 0.02$) due principally to a higher frequency of haplotype 122, and a lower incidence of 211, in Italians. Slav haplotype frequencies are more similar to Greeks than Italians. No statistically significant regional differences exist within either Italy or Greece, though the frequency of 211 rises from 6% in the north-central of Italy to 37% in the south.

Haplotype 21A1, 92R7 and pYαI. All four expected haplotypes were observed in Italians and Slavs, but not in Greeks who lack haplotype 122 (Table 3). Allele 1 of p21A1 was preferentially associated with a Y chromosome containing allele 1 of 92R7

TABLE 2. Distribution of Y-specific three loci haplotypes (12f, 92R7 and YαI) in European and Asian populations

Population and birthplace	Haplotype 111 No.	111 %	122 No.	122 %	211 No.	211 %	222 No.	222 %	Total
Italy									
North-central	14	45.2	15	48.4	2	6.4	0	0.0	31
South	4	21.0	8	42.1	7	36.9	0	0.0	19
Sicily	9	42.9	7	33.3	5	23.8	0	0.0	21
Total	28	37.8	32	43.3	14	18.9	0	0.0	74
Greece									
Mainland	20	46.5	10	23.3	13	30.2	0	0.0	43
Islands	6	40.0	2	13.3	7	46.7	0	0.0	15
Total	26	44.8	12	20.7	20	34.5	0	0.0	58
Slavs	8	47.1	5	29.4	4	23.5	0	0.0	17
Total Europe	62	41.6	49	32.9	38	25.5	0	0.0	149
Asians	23	95.8	0	0.0	1	4.2	0	0.0	24

Locus order—12f, 92R7 and YαI.

TABLE 3. Distribution of Y-specific three loci haplotypes (21A1, 92R7 and YαI) in European and Asian populations

Population and birthplace	Haplotype 111 No.	111 %	122 No.	122 %	211 No.	211 %	222 No.	222 %	Total
Italy									
North-central	7	22.6	2	6.5	9	29.0	13	41.9	31
South	5	26.3	0	0.0	6	31.6	8	42.1	19
Sicily	7	33.3	1	4.8	7	33.3	6	28.6	21
Total	20	27.0	3	4.1	22	29.7	29	39.2	74
Greece									
Mainland	13	28.9	0	0.0	22	48.9	10	22.2	45
Islands	5	31.2	0	0.0	9	56.3	2	12.5	16
Total	18	29.5	0	0.0	31	50.8	12	19.7	61
Slavs	5	45.4	1	9.1	2	18.2	3	27.3	11
Total Europe	43	29.5	4	2.7	55	37.7	44	30.1	146
Asians	3	13.0	0	0.0	20	87.9	0	0.0	23

Locus order—21A1, 92R7 and YαI.

and allele 1 of pYαI. Allele 2 of p21A1, however, was seen on both chromosomes, but more frequently on 11 rather than 22 chromosomes in Greeks. The reverse is observed in Italians and Slavs.

The Asian sample is significantly different from all three European populations (vs. Greeks $P < 0.01$; vs. Italians $P < 0.001$; vs. Slavs $P < 0.01$), principally because of the absence of haplotypes 122 and 222, and their extreme frequency of 211. Italians have a low frequency of 111 and a high frequency of 122 compared to Greeks, and these differences are significant ($P < 0.02$). No regional differences of significance were seen among either Greeks or Italians.

Haplotype 12f, 21A1, 92R7 and pYαI. The most common four-loci haplotype in Italians is 1222 (followed by 1211 and 1111), while in Greeks the most common is 1211 (followed by 1222 and 2211) (Table 4). Only three haplotypes are seen in Asians, with one, 1211, markedly predominant. From the data in Table 4 it appears that p21A1 must have undergone recurrent mutation in European populations. Both alleles at this locus are found on all three haplotypes seen in Table 2, though allele 1 is not on 122 in Greeks and allele 2 is not found with 211 in Slavs. This phenomenon makes haplotype distributions involving p21A1 alleles difficult to interpret.

DISCUSSION

No other population study of Y chromosome polymorphisms has included all of the four RFLPs used in our study, and the data available for 92R7 and pYαI are scanty, with only a handful of subjects typed from the major population groups, except Caucasians

TABLE 4. Y-specific four loci haplotypes, (12f, 21A1, 92R7 and YαI) in Europeans and Asians

	Haplotype												
	1111		1122		1211		1222		2111		2211		
Population	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	Total
Italy	14	18.9	3	4.1	14	18.9	29	39.2	6	8.1	8	20.8	74
Greece	9	15.5	0	0.0	17	29.3	12	20.7	8	13.8	12	20.7	58
Slavs	2	18.2	1	9.1	2	18.2	3	27.3	3	27.3	0	0.0	11
Total Europe	25	17.5	4	2.8	33	23.1	44	30.8	17	11.0	20	14.0	143
Asians	2	8.7	0	0.0	20	87.0	0	0.0	1	4.3	0	0.0	23

Locus order—12f, 21A1, 92R7 and YαI.

(Persichetti et al., 1992; Mathias et al., 1994). Accordingly, the present data are discussed first in terms of allele frequencies at each locus to enable comparisons with other data, and then in haplotype combinations which are usually more informative. The origins and structural basis of some of the other Y markers that have been used in other studies are unclear. For example, multiple origins of alleles occurred at the pDP31, p49a/p49f, and DYS19 loci (Spurdle et al., 1994a; Spurdle and Jenkins, 1994; Hammer and Horai, 1995). This mechanism seems also to be true for p21A1 in our sample, as well as in others (Spurdle et al., 1994a), and thus p21A1's usefulness as an indicator of population affinities is minimal. Alleles at the other three loci examined in this study may have arisen only once in evolution, and accordingly should be more informative about relationships among major population groups. The repetitive structure of the alphoid sequence detected by pYαI, however, presents multiple targets for base substitutions; therefore, the possibility of recurrent mutation cannot be excluded at this locus (Oakey and Tyler-Smith, 1990).

The frequency of 92R7 allele 1 reported in Caucasians (60%) by Mathias et al. (1994) is more similar to that seen in Italians than Greeks, and the monomorphism of Asians at this locus is consistent with their findings based on a very small sample. The frequency of pYαI alleles varies markedly across the major populations (Persichetti et al., 1992; Mathias et al., 1994) and also among Caucasian groups. The various regional Italian samples of Persichetti et al. (1992) and our data exhibit considerable variation (allele 1 ranging from 52% in north-central Italy to 87% in north Sardinia). Our larger Italian sample shows a much lower frequency of allele 1 (57%) than the smaller sample from mainland Italy (80%) obtained by Persichetti et al. (1992), and the frequency we observed is much closer to that seen in the Caucasian sample of Mathias et al. (1994). The high frequency of pYαI allele 1 in Greeks and Slavs resembles that of other populations of the Mediterranean Basin, such as Egyptians (85%) and Sardinians. They are unlike the English frequency (33%) which is the lowest among Europeans. The available data suggest a cline in the frequency of the pYαI 6.0 kb allele across Europe, from a low in the north west (England) to a maximum in the south-east (Greece). The present study confirms this locus is monomorphic in Asians.

Slavs exhibit one of the lowest frequencies of the mutant allele at p21A1 with only Polynesians having a lower incidence (Spurdle et al., 1994b). Most populations exhibit frequencies of >60% for allele 2 (Spurdle and Jenkins, 1993b). That it is polymorphic in all populations, combined with evidence of multiple origins of the variant allele makes p21A1 relatively uninformative for investigation of either population affinities or Y chromosome evolution.

p12f 8 kb allele: Its origin and dispersion

The p12f/*Taq*I 8 kb allele is absent in Africans (Casanova et al., 1985; Brega et al., 1987; Semino et al., 1995) and in Amerindians (Torroni et al., 1994). Semino et al. (1995) report it absent in a sample of 263 Orientals (from China, Malaya and Indonesia), which strongly suggests that the single 8 kb allele in our Asian sample is a result of recent admixture. Among Caucasoids, however, the 8 kb allele frequency varies considerably. Lower frequencies (<15% generally) are found in northwestern and central Euro-

peans, such as the French (Casanova et al., 1985) and Czechs (Santachiara Benerecetti et al., 1993), and also historically relatively isolated groups, such as the Basques (Santachiara Benerecetti et al., 1995). In populations bordering the Mediterranean Sea the frequency is considerably higher, usually >30%, (Casanova et al., 1985; Brega et al., 1987; Mitchell et al., 1993; Santachiara Benerecetti et al., 1993) with the maximum level seen in Lebanese and Jewish groups. Slavs display a frequency consistent with their geographic position between the Mediterranean and eastern Europe.

The present data on p12f strongly suggest that the polymorphism arose from a deletion event that may have only occurred once in human evolution, in a Caucasian male who possessed, among other characteristics, allele 1 at the 92R7 locus and the 6.0 kb alphoid units at DYZ3. To confirm the uniqueness of the 8 kb allele at DYS11 would require the probing of males carrying this 8 kb allele for other unique Y-specific markers. Of particular value would be the Y *Alu* polymorphism or YAP (DYS 287) (Hammer, 1994). Persichetti et al. (1992) found that alleles at the YAP and pYαI loci exhibited highly significant linkage disequilibrium in their five Caucasoid populations. The YAP+ allele frequency is approximately 10% in Europeans and the association between both p12f and YAP alleles and between 92R7 and YAP alleles needs to be examined. In a preliminary analysis of the relation between YAP and p12f we found that all Y chromosomes carrying the p12f 8 kb allele were also YAP− (Mitchell and Hammer, unpublished results). These limited data permit the identification of the Y-chromosome haplotype of p12f/*Taq*I/8 kb; 92R7 *Hin*dIII 4.6 kb; pYαI/*Hin*dIII 6.0 kb+; YAP− as potentially distinctive of Caucasoids. The data on these three loci in this study also indicate that, among Caucasoids, at least two other Y-specific haplotypes exist: p12f 10 kb/92R7 4.6 kb/pYαI 6.0 kb+ and p12f 10 kb/92R7 6.7 kb/pYαI 6.0 kb−.

European populations bordering the Mediterranean Sea may have many African genes (Gonçalves and Lavinha, 1994). This is possibly reflected in the high frequency of the 92R7/pYαI haplotype 11, compared to populations elsewhere in Europe. African admixture, however, cannot explain the high frequency of the p12f/92R7/pYαI haplotype, 211, (that is also YAP−) in the same populations. Two questions are posed by the distribution of this distinctively Caucasian haplotype: 1) where is the focus, and the likely origin of this haplotype, which arose after *Homo sapiens* had left Africa? and 2) how did it spread to attain its present distribution?

The current distribution of haplotype 211 suggests that the focus lies in the Near East (e.g., Lebanon) but it is possible that its true focus is among groups further east, in Syria and Iraq, for which there are, as yet, no data, except for Jewish groups living in this region, among whom it is high (Ritte et al., 1993). The spread of 211 from its epicentre may have been a result of migration of farmers from the Middle East during the Neolithic transition, a mechanism Gonçalves and Lavinha (1994) have suggested to explain the relatively high frequency in Portuguese of the '*low*' allele of the Y pseudoautosomal marker, XY275 or *PABY*. However, if brought to Europe by farmers, one would not expect to find a high frequency of the 211 haplotype restricted to those European (and North African) populations that are adjacent to the prehistoric sea routes of the Mediterranean Basin, and yet this is the distribution pattern seen. One cannot rule out the possibility that this Y-specific haplotype was spread by sea-going peoples, and not cultivators, who travelled from the Near East, into parts of north Africa and Europe that border the Mediterranean. The ancient Phoenicians are a strong candidate in this regard. They inhabited the east Mediterranean coast nearly 4,000 years ago and established trade routes throughout almost every corner of the Mediterranean Basin. Whether they were indigenous to the area or came from the Aegean is still disputed. Considerably more data on Y-chromosome specific variation are required to resolve these issues.

If the evolution of the p12f/*Taq*I 8 kb allele, or haplotype 211, proves to be a unique evolutionary event then Y chromosomes containing it should prove invaluable in measuring Caucasoid male gene flow, particularly where southern Europeans/

north Africans are involved. Torroni et al. (1994) found evidence of Caucasoid admixture in Native Americans of Mexico based on p49a haplotypes but no such evidence using p12f. Semino et al. (1995) report significant disequilibrium between alleles at p49a/f and p12f such that 85% of all p12f/8 kb alleles they found were associated with three p49a/f haplotypes, (Ht no. 8, Ht no. 7 and Ht no. 24 respectively). They further suggest that all Ht no. 24 and Ht no. 27 chromosomes are derived from a Ht no. 8 chromosome that carried the p12f/8 kb fragment. The frequency of the p12f 8 kb allele in Spain (the colonisers of South and Central America) is currently unknown but if found to be similar to those in other Mediterranean groups it could prove useful in this regard. We caution that Basques in northern Spain have a very low frequency of the allele.

Y-specific haplotypes and diversity

Using a number of polymorphisms Oakey and Tyler-Smith (1990) and later Mathias et al. (1994) found that the Y chromosomes from a sample of men of European origin could be traced to one of two ancestors. The ancestor of what they labelled group 1 chromosomes lacked the 6.0 kb units with two *Hin*dIII sites recognised by pYαI and carried the 6.7 kb allele at 92R7 (among other features), whereas the ancestor of their group 2 contained the additional sites for *Hin*dIII and the 4.6 kb allele at 92R7. The present study demonstrates that these group 2 chromosomes can be subdivided on the basis of the p12f polymorphism. One group contains the 6.0 kb alphoid units at DYZ3, the 4.6 kb allele at 92R7 and the 10 kb allele at DYS11 (group 2), whereas the other has the 6.0 kb units at DYZ3, the 4.6 kb allele at 92R7, and the 8 kb allele at DYS11 (and is also apparently YAP−). This latter chromosome would lie within the 'Mediterranean' group recently proposed by Jobling and Tyler-Smith (1995), which is distinguished from other Y haplotypes by the p12f allele. This 'Mediterranean' Y chromosome is more recent than group 2 and, based on present data, appears to be restricted to Caucasians. Group 2 chromosomes are found in both Caucasian and non Caucasian groups.

Caucasians show considerable diversity of Y chromosomes when compared to Africans, which is the reverse of findings with respect to the levels of mtDNA and autosomal sequence diversity. However, the differences in the sheer volume of data available for the latter, compared to that for the male-specific chromosome, stress caution in any interpretation of this observation. Caucasians exhibit high frequencies of two Y chromosomes that are apparently unique to them (group 1 and the 'Mediterranean') in addition to relatively high frequencies of group 2 and even group 4 (characterised by YAP+) chromosomes. It has been proposed that group 2 is the oldest Y chromosome in humans, and that separate mutations on this particular chromosome produced the additional three types found in Europeans today (Jobling and Tyler-Smith, 1995; Tyler-Smith and Hammer, 1995). Modern Africans possess one of two types of Y chromosome while in Chinese and Vietnamese there is evidence of only one Y chromosome so far (group 2). However, both the number of Asian populations reported in the literature for these unique Y-specific alleles and the number in each sample are extremely small. A larger Asian data base may alter current conclusions substantially.

This study has shown that among Europeans complete disequilibrium exists between alleles at the 92R7 and pYαI loci. Furthermore, the Y-specific haplotype comprising the 6.7 kb allele at 92R7 and 6.0 kb− allele at pYαI has its highest frequency in this group, and is most prevalent in northwest Europe. Our data also confirm that the 8 kb allele at the p12f locus probably arose only once, as a deletion event, and relatively recently as it generated a Y-specific haplotype apparently restricted to Caucasian groups. Fuller understanding of the geographical origin and subsequent spread of these two Y haplotypes may result from the scoring of highly polymorphic microsatellites recently discovered on the Y-specific chromosome (Roewer et al., 1992, 1996; Santos et al., 1993; Mathias et al., 1994; Gomolkaert et al., 1994; Muller et al., 1994; Ciminelli et al., 1996) in those Y chromosomes carrying these 'unique' or rare variants. Haplotype associations of these simple

repeat markers with sequence polymorphisms could prove highly informative in both tracing the direction and time of spread of the p12f deletion chromosome throughout the Mediterranean basin.

Further significant advances in the understanding of both Y chromosome evolution and the role of male gene flow in explaining human diversity require at least two developments. First, many more populations must be examined, and in particular samples from the numerous indigenous groups of Asia and the Americas are needed to avoid a Eurocentric bias in interpretation of the findings. To date, inadequate numbers of both populations and individuals from each group have been examined for Y-specific polymorphisms, with the possible exceptions of p49a/f, which is subject to recurrent mutation, duplication and re-arrangements, and YAP. Perhaps more importantly, an agreed minimum set of markers, with newly discovered ones added to the set as they prove suitable, should be examined in all samples, even if this involves more than one laboratory. Analysis could then be undertaken and presented at the level of haplotypes which, given the non recombining behaviour of the Y chromosome, is so much more informative. The Y Chromosome Consortium is working towards meeting both these goals.

ACKNOWLEDGMENTS

We gratefully acknowledge C. Tyler-Smith, CRC Chromosome Molecular Biology Group, Department of Biochemistry, University of Oxford, UK, for his gift of probes 92R7 and pYαI and control DNAs, J. Weissenbach, Institut Pasteur, Paris, France, for probe 12f, and H. Cooke, MRC Human Genetics Unit, Edinburgh, UK, for probe p21A1.

LITERATURE CITED

Brega A, Scozzari R, Maccioni L, Jodice C, Wallace DC, Bianco I, Cao I, and Santachiara Benerecetti AS (1986) Mitochondrial DNA polymorphisms in Italy. I Population data from Sardinia and Rome. Ann. Hum. Genet. *50:*327–338.

Brega A, Torroni A, Semino O, Maccioni I, Casanova M, Scozzari R, Fellous M, and Santachiara-Benerecetti AS (1987) The p12f2/TaqI Y-specific polymorphism in three groups of Italians and in a sample of Senegalese. Gene Geogr. *1:*201–206.

Cann RL, Stoneking M, and Wilson AC (1987) Mitochondrial DNA and human evolution. Nature *325:*321–336.

Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Purello M, Fiori G, and Siniscalco M (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science *230:*1403–1406.

Charlesworth B (1991) The evolution of sex chromosomes. Science *251:*1030–1033.

Ciminelli BM, Pompei F, Malaspina P, Hammer M, Persichetti F, Pignatti PF, Palena A, Anagnou N, Guanti G, Jodice C, Terrenato L, and Novelletto A (1996) Recurrent simple tandem repeat mutations during human Y-chromosome radiation in Caucasian subpopulations. J. Mol. Evol. *41:*966–973.

Cooke HJ, Brown WRA, and Rappold GRA (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. Nature *317:*687–692.

Dorit RL, Akashi H, and Gilbert W (1995) Absence of polymorphism at the ZFY locus on the human Y chromosome. Science *268:*1183–1185.

Freije D, Helms C, Watson MS, and Donis-Keller H (1992) Identification of a second pseudoautosomal region near the Xq and Yq telomeres. Science *258:* 1784–1787.

Gomolka M, Hundrieser J, Nurnberg P, Roewer L, Epplen JT, and Epplen C (1994) Selected di- and tetranucleotide microsatellites from chromosomes 7, 12, 14, and Y in various Eurasian populations. Hum. Genet. *93:*592–596.

Gonçalves J, and Lavinha J (1994) The Y-associated XY275G (low) allele is common among the Portuguese. Am. J. Hum. Genet. *55:*583–585.

Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol. Biol. Evol. *11:*749–761.

Hammer MF (1995) A recent common ancestry for human Y chromosomes. Nature *378:*376–378.

Hammer MF, and Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. Am. J. Hum. Genet. *56:*951–962.

Jakubiczka S, Arnemann J, Cooke HJ, Krawczak M, and Schmidtke J (1989) A search for restriction length polymorphism on the human Y chromosome. Hum. Genet. *84:*86–88.

Jobling MA (1994) A survey of long-range DNA polymorphisms on the human Y chromosome. Hum. Mol. Genet. *3:*107–114.

Jobling MA, and Tyler-Smith C (1995) Fathers and sons: The Y chromosome and human evolution. Trends Genet. *11:*449–456.

Malaspina P, Persichetti F, Novelletto A, Jodice C, Terrenato L, Wolfe J, Ferraro M, and Prantera G (1990) The human Y chromosome shows a low level of DNA polymorphism. Ann. Hum. Genet. *54:*297–305.

Mathias N, Bayes M, and Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. Hum. Mol. Genet. *3:*115–123.

Maynard Smith J (1990) The Y of human relationships. Nature *344:*591–592.

Mitchell RJ, Earl L, and Williams JW (1993) Two Y-chromosome-specific restriction fragment length polymorphisms (DYS11 and DYZ8) in Italian and Greek migrants to Australia. Hum. Biol. *65:*387–399.

Muller S, Gomolka M, and Walter H (1994) The Y-specific SSLP of the locus DYS19 in four different European samples. Hum. Hered. *44:*298–300.

Ngo KY, and Lucotte G (1986) Strategies de recherche des polymorphismes de restriction des sequences du chromosome Y. Ann. Genet. *29:*88–92 (in French with English summary).

Oakey R, and Tyler-Smith C (1990) Y chromosome DNA haplotyping suggests that most European and Asian men are descendant from one of two males. Genomics *7:*325–330.

Persichetti F, Blasi P, Hammer M, Malaspina P, Jodice C, Terrenato L, and Novelletto A (1992) Disequilibrium of multiple DNA markers on the human Y chromosome. Ann. Hum. Genet. *56:*303–310.

Ritte U, Neufield E, Broit M, Shavit D, and Motro U (1993) The difference among Jewish communities—maternal and paternal contributions. J. Mol. Evol. *37:*435–440.

Roewer L, Arnemann J, Spurr NK, Grzeschik K-H, and Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum. Genet. *89:*389–394.

Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, and de Knijff P (1996) Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. Hum. Molec. Genet. *5:*1029–1033.

Sambrook J, Fritsch EF, and Maniatis T (1989) Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Santachiara Benerecetti AS, Semino O, Passarino G, Torroni A, Brdicka R, Fellous M, and Modiano G (1993) The common Near-Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. Ann. Hum. Genet. *57:*55–64.

Santachiara Benerecetti AS, Semino O, Passarino G, Bertranpetit J, and Fellous M (1995) The genetic peculiarity of the Basques shown by some Y-specific polymorphisms. Am. J. Hum. Genet. *57*A948.

Santos FR, Pena SDJ, and Epplen JT (1993) Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. Hum. Genet. *90:*655–656.

Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, and Cavalli-Sforza LL (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum. Mol. Genet. *3:*2159–2161.

Semino O, Passarino G, Liu A, Brega A, Fellous M, and Santachiara Benerecetti AS (1995) Three Y-specific polymorphisms in populations of different ethnic and geographic origin. Y Chromosome Consortium Newsletter *2:*5–6. Department of EEB, University of Arizona, Tucson.

Spurdle A, and Jenkins T (1992a) The search for Y chromosome polymorphism is extended to negroids. Hum. Mol. Genet. *1:*169–170.

Spurdle AB, and Jenkins T (1992b) The Y chromosome as a tool for studying human evolution. Curr. Opin. Genet. Dev. *2:*487–491.

Spurdle AB and Jenkins T (1993a) Complex polymorphisms are revealed by Y chromosome probe 49a with *Bgl*II, *Hind*III, *Pst*I and *Sst*I. Ann. Hum. Genet. *57:*41–53.

Spurdle A and Jenkins T (1993b) The Y-specific p21A1/TaqI polymorphisms occur in four major population groups. Hum. Hered. *43:*31–34.

Spurdle AB and Jenkins T (1994) The Y-specific pDP31 rearrangement polymorphism in southern African populations. Hum Hered. *44:*261–265.

Spurdle A, Hammer M, and Jenkins T (1994a) The Y Alu polymorphism in southern African populations, and its relationship to other Y specific polymorphisms. Am. J. Hum. Genet. *54:*319–330.

Spurdle AB, Woodfield DG, Hammer MF, and Jenkins T (1994b) The genetic affinity of Polynesians: Evidence from Y chromosome polymorphisms. Ann. Hum. Genet. *58:*251–263.

Torroni A, Semino O, Scozzari R, Sirugo G, Spedini G, Abbas N, Fellous M, and Santachiara Benerecetti AS (1990) Y chromosome DNA polymorphism in human populations: Differences between Caucasoids and Africans detected by 49a and 49f probes. Ann. Hum. Genet. *54:*287–296.

Torroni A, Chen Y-S, Semino O, Santachiara-Benerecetti AS, Scott CR, Lott MT, Winter M, and Wallace DC (1994) mtDNA and Y-chromosome polymorphisms in four native American populations from southern Mexico. Am. J. Hum. Genet. *54:*303–318.

Tyler-Smith C, and Brown WRA (1987) Structure of the major block of alphoid satellite DNA on the human Y chromosome. J. Mol. Biol. *195:*457–470.

Tyler-Smith C, and Hammer M (1995) The relationships of Y chromosomes determined using five RFLPs and the YAP element. Y Chromosome Newsletter *2:*4–5.

Whitfield LS, Sulston JE, and Goodfellow PN (1995) Sequence variation of the human Y chromosome. Nature *378:*379–380.